POLYSATURATED FATTY ACID (PUFA) ELONGASE FROM CAENORHABDITIS ELEGANS

Insert

The present invention relates to polyunsaturated fatty acid (PUFA) elongases. More specifically, the invention relates to a DNA sequence from *C. elegans* encoding a PUFA elongase.

Unsaturated fatty acids are essential components required for normal cellular function, being involved in a diverse number of roles ranging from membrane fluidity to acting as signal molecules (Gill, I., Valivety, R. (1997). *Trends Biotechnol.* **15**, 401-409; Broun, P., *et al* (1999) *Ann. Rev. Nutr.* **19**, 197-216). In particular, the class of fatty acids known as the polyunsaturated fatty acids (PUFAs) has attracted considerable interest as pharmaceutical and nutraceutical compounds (Broun *supra*; Horrobin, D. F. (1990) *Reviews in Contemp Pharmacotherpy* **1**, 1-45).

The synthesis of PUFAs i.e. fatty acids of 18 carbons or more in length and containing two or more double bonds, is thought to be catalyzed in a variety of organisms by a specific fatty acid elongase enzyme. This elongase is responsible for the addition of 2 carbon units to an 18 carbon PUFA, resulting in a 20 carbon fatty acid. An example of this reaction is the elongation of γ-linolenic acid (GLA; $18:3\Delta^{6,9,12}$) to di-homo-γ-linolenic acid (DHGLA; $20:3\Delta^{8,11,14}$) in which the tri-unsaturated 18 carbon fatty acid is elongated by the addition of a two carbon unit to yield the tri-unsaturated 20 carbon fatty acid. Since there is considerable interest in the production of long chain PUFAs of more than 18 carbons in chain length, for example arachidonic acid and eicosapentanoic acid, the identification of this enzyme is of both academic and commercial interest.

At present, there are no examples of identified cloned genes encoding PUFA elongases, though a number of genes encoding enzymes likely to be involved in other aspects of lipid synthesis have been identified. For example, an *Arabidopsis* gene (FAE1) has been shown to be required for the synthesis of very long chain monounsaturated fatty acids (such as erucic acid; $20:1\Delta^{11}$) (James, D. W. *et al*, (1995) *Plant Cell* 7, 309-319). However, it is clear that this enzyme does not recognize di- and tri-unsaturated 18 carbon fatty acids, for example, linoleic acid, $18:2\Delta^{9,12}$ or α-linolenic acid, $18:3\Delta^{9,12,15}$ respectively, as substrates,

and is therefore not involved in the synthesis of long chain PUFAs (Millar & Kunst (1997), *Plant Journal* **12**, 121-131). This in itself is not surprising, since, of the plant kingdom, only a very few lower plant species, such as the moss *Physcomicotrella patens* (Girke *et al.*, (1998), *Plant J,* **15**: 39-48); are capable of synthesising long chain PUFAs, and therefore *Arabidopsis* would not be expected to contain any such enzymes (Napier *et al.* (1997), Biochem J, **328**: 717-720; Napier *et al.*, (1999) *Trends in Plant Sci* **4**, 2-5).

A schematic diagram representing a generalized pathway for the product of PUFAs is shown in Figure 1. Biochemical characterisation of mammalian elongation systems (most notably from liver microsomes) has indicated that a mammalian elongase consists of four subunits, made up of a condensing enzyme, a β-ketoreductase, a dehydrase and an enoyl reductase (reviewed in Cinti, D. L., *et al* (1992) *Prog. Lipid Res.* **31**, 1-51). The *Arabidopsis FAE1* gene product encodes a polypeptide of 56kDa, which shows very limited homology to condensing enzymes such as chalcone synthase and stillbene synthase (James, D. W. *supra*). Although *FAE1* is normally only expressed in seed tissues, ectopic expression in non-seed tissue (or heterologously in yeast) revealed that *FAE1* could direct the synthesis of erucic acid (Millar, A. A., Kunst, L. (1997) *Plant J.* **12,** 121-131).

Three fatty acid elongase activities have been characterised from the yeast *S. cerevisiae.* Again, this organism does not synthesis PUFAs, and therefore does not contain genes encoding a PUFA elongase. One gene ELO1, was identified on the basis of a screen to isolate mutants defective in elongation of 14 carbon (i.e. medium) chain saturated fatty acids (Toke & Martin (1996) *J Biol Chem* **271**, 18413-18422). Complementation of *elo1* mutants restored viability, and the ELO1 gene product was shown to encode a polypeptide which was responsible for the specific elongation of 14:0 fatty acids to 16:0 fatty acids.

Two related genes were also detected in the genome of *S. cerevisiae*, and their function determined by disruption. These two genes, subsequently named ELO2 and ELO3, were shown to be involved in the elongation of the very long chain saturated fatty acids found in sphingolipid molecules (Oh *et al* (1997), *J. Biol Chem* **272**, 17376-17384). In particular, ELO2 was required for elongation of fatty acids up to 24 carbons, and ELO3 was required for elongation of the 24 carbon fatty acid to 26 carbons. However, neither gene was

essential for viability. Examination of the these three fatty acid elongases revealed the presence of a conserved "histidine box" motif (Shanklin *et al.*, (1994), *Biochemistry*, **33**, 12787-12794) (His-X-X-His-His, where X is any amino acid) towards the centre of the polypeptide sequences. Importantly, there was no detectable homology between the yeast elongases (ELO1,2,3) and the plant very long chain mono-unsaturated fatty acid elongase (FAE1) (Oh *et al, supra*).

In order to identify genes encoding PUFA elongases, it is necessary to study systems in which the synthesis of PUFAs is well documented; a good example of this is the model animal system *C. elegans*, a small free-living worm (Tanaka *et al.*, (1996), *Lipids* **31**, 1173-1178). *C. elegans*, like most other animals, and in contrast to higher plants, synthesises PUFAs such as arachidonic acid (AA; 20:4 $\Delta^{5,8,11,14}$) as precursors to a class of molecules known as the eicosanoids, which in turn serve as precursors for compounds such as prostaglandins and leucotrienes (Horrobin, (1990), *Reviews in Contemp Pharmacotherpy*, **1**:1-45). The presence of AA and other long chain polyunsaturated fatty acids in *C. elegans* is well documented (Tanaka *et al.*, (1996), *Lipids* **31**, 1173-1178). The complete sequence of the nematode's genome is now publicly available *(The C. elegans consortium, 1998, Science* ***282****, 2012-2018: Database at http://www.sanger.ac.uk/Projects/C_elgans/blast_server.shtml).*

An object of the invention is to provide an isolated PUFA elongase.

Using the above-mentioned *C. elegans* genomic sequence, together with suitable search strings, the inventors identified eight related putative open reading frames (ORFs) encoding for PUFA elongases. A number of different search criteria were applied to identify a number of (ORFs) which were likely to encode polypeptides with fatty acid elongase activities. These ORFs were then subject to functional characterisation by heterologous expression in yeast, allowing the identification of a PUFA elongase.

Accordingly, a first aspect of the invention provides an isolated polypeptide comprising a functional long chain polyunsaturated fatty acid (PUFA) elongase i.e. the polypeptide has the function of extending the chain length of an 18 carbon PUFA to 20 carbons in length.

This polypeptide can be used to elevate PUFA levels in animals, thereby providing a ready source of PUFAs.

The polypeptide may be from a eukaryote.

The polypeptide may comprise at least a portion of the amino acid shown in SEQ ID. 15, or variants thereof.

For the purposes of the present application, the term "variant" in relation to a certain sequence means a protein or polypeptide which is derived from the sequence through the insertion or deletion of one or more amino acid residues or the substitution of one or more animo acid residues with amino acid residues having similar properties, e.g. the replacement of a polar amino acid residue with another polar amino acid residue, or the replacement of a non-polar amino acid residue with another non-polar amino acid residue. In all cases, variants must have an elongase function as defined herein.

A second aspect of the invention provides a polypeptide having at least 60 % homology to a polypeptide according to a first aspect of the invention. The polypeptide may have at least 80%, or as much as 90% or more homology to a polypeptide according to a first aspect of the invention.

The polypeptide according to either aspect of the invention may include a sequence motif responsible for Endoplasmic Reticulum (ER) - retention. This allows the polypeptide to be specifically located or targeted to the ER of a cell.

The polypeptide may also be able to elongate palmitoleic acid (PA; 16:1Δ^9) to vacceric acid (VA; 18:1Δ^{11}). Thus, the polypeptide is also capable of elongation of a Δ^9- monounsaturated 16C fatty acid.

Preferably, the polypeptide is from an animal, more preferably, the animal is an invertebrate such as a worm. Where the animal is a worm, it is preferably *C. elegans*. Alternatively, the animal is a vertebrate, preferably a mammal such as a human, rat or mouse.

A third aspect of the invention provides an isolated DNA sequence, preferably a cDNA sequence, encoding a polypeptide according to a first or second aspect of the invention. This DNA sequence may be used to engineer transgenic organisms.

Preferably, the DNA sequence comprises the sequence shown in SEQ ID NO: 7 or variants of that sequence due, for example, to base substitutions, deletions, and/or additions.

A fourth aspect of the invention provides an engineered organism, such as a transgenic animal, engineered to express a polypeptide according to a first or second aspect of the invention. The engineered organism may be engineered to express elevated levels of the polypeptide, thereby providing a supply of polypeptide at a reduced cost as a reduced number of organisms need be used.

Preferably, the engineered organism is a mammal such as a rat, mouse or monkey.

A fifth aspect of the invention provides an engineered organism containing a synthetic pathway for the production of a polypeptide according to a first or second aspect of the invention. This has the advantage of allowing greater control over the production of PUFAs by the pathway by an organism.

The pathway may include Δ^5-fatty acid desaturase, and/or Δ^6-fatty acid desaturase.

The engineered organism according to a fourth or fifth aspect of the invention may be a lower eukaryote, such as yeast. Alternatively, the transgenic organism may be a fish.

A sixth aspect of the invention provides a transgenic plant engineered to express a polypeptide according to a first aspect of the invention.

A seventh aspect of the invention provides a transgenic plant containing a DNA sequence according to a third aspect of the invention.

An eighth aspect of the invention provides a method of producing a PUFA comprising carrying out an elongase reaction catalysed by a polypeptide according to a first or second aspect of the invention.

The PUFA may be di-homo-gamma-linoleic acid ($20:3\Delta^{8,11,14}$), arachidonic acid ($20:4\Delta^{5,8,11,14}$), eicosapentanoic acid ($20:5\Delta^{5,8,11,14,17}$), docosatrienoic acid ($22:3\Delta^{3,16,19}$), docosatetraenoic acid ($22:4\Delta^{7,10,13,16}$), docosapentaenoic acid ($22:5\Delta^{7,10,13,16,19}$) or docosahexaenoic acid ($22:6\Delta^{4,7,10,13,16,19}$).

The PUFA may be a 24 carbon fatty acid with at least 4 double bonds.

A ninth aspect of the invention provides a PUFA produced by a method according to an eighth aspect of the invention.

The PUFA may be used in foodstuffs, dietary supplements or pharmaceutical compositions.

A tenth aspect of the invention provides a foodstuff comprising a PUFA according to a fifth aspect of the invention. The foodstuff can be fed to an animal.

An eleventh aspect of the invention provides a dietary supplement comprising a PUFA according to a fifth aspect of the invention. The dietary supplement can be supplied to an animal to augment its PUFA levels.

An twelfth aspect of the invention provides a pharmaceutical composition comprising a polypeptide according to a first or second aspect of the invention or a PUFA according to a ninth aspect of the invention.

Preferably, the pharmaceutical composition comprises a pharmaceutically-acceptable diluent, carrier, excipient or extender. This allows the composition to be supplied in a form which best suits the pharmaceutical application in question. For example, a topical application would preferably be a cream or lotion, whereas if the composition was to be ingested a different form would be more suitable.

A thirteenth aspect of the invention provides a method of treatment of an animal, such as a mammal, or a plant, comprising supplying to the animal or plant a DNA sequence according to a third aspect of the invention, a foodstuff according to a tenth aspect of the invention, a dietary supplement according to an eleventh aspect of the invention, a pharmaceutical composition according to a twelfth aspect of the invention or a PUFA according to a ninth aspect of the invention.

Preferably, the mammal is a human.

The invention will now be further described, by way of example only, with reference to SEQ ID1 to 16, and Figures 2 to 11, in which;

SEQ ID1 to 8 show the putative ORFs encoding PUFA elongases A to H respectively; and

SEQ ID9 to 16 show the deduced amino acid sequences of the putative ORFs of SEQ ID NO: 1 to 8 respectively; and

Figures 2 to 9 show hydrophobicity plots for each of PUFA elongases A to H respectively.

Figure 10 shows an amino acid sequence line-up comparing the *C. elegans* ORF F56H11.4 (Z68749) with related sequences.

Figure 11 shows chromatograms of fatty acid methyl esters from transformed yeast.

Introduction to general strategy

Initially the *C. elegans* databases were searched for any sequences which showed low levels of homology to yeast ELO genes (*ELO2* and *ELO3*) using the TBLASTN programme. A similar search was carried out using short (20 to 50 amino acid) stretches of ELO genes which were conserved amongst the three ELO polypeptide sequences. *C. elegans* sequences which were identified by this method were then used themselves as search probes, to identify any related *C. elegans* genes which the initial search with the yeast sequences failed to identify. This was necessary because the level of homology between the yeast ELO genes

and any worm genes is always low (see BLAST scores later). To allow for a more sensitive search of worm sequences, a novel approach was adopted to circumvent the major drawback with searches using the BLAST programmes, namely that the search string (i.e. the input search motif) must be longer than 15 characters for the algorithm to work. Thus, if it was desired to search for a short motif (like a histidine box), then the BLAST programme would not be capable of doing this. A complete list of all the predicted ORFs present in the *C. elegans* genome exists as a database called Wormpep, which is freely available from the Sanger WWW site (http://www.sanger.ac.uk/Projects/C_elegans/webace_front_end.shtml). The latest version of Wormpep was down loaded to the hard disc of a Pentium PC, and re-formatted as a Microsoft Word6 document, resulting in a document of about 3,500 pages. This was then searched using the "Search & Replace" function of Word6, which also allows for the introduction of "wildcard" characters into the search motif. So, for example, it is possible to search both for the short text string HPGG, which would identify any predicted worm ORF present in the Wormpep 3,500 page document containing this motif, or alternatively search with HPGX (where X is a wild card character). Clearly, such (manual) searches of a 3,500 page document are extremely time-consuming and demanding, also requiring visual inspection of each and every identified ORF. For example, searching with a motif such as HXXHH identifies in excess of 300 different ORFs. However, by using a number of different short search strings (as outlined below), and combining these with other methods for identifying putative elongase enzymes, a number of candidate ORFs have been identified.

sub a4

Database search using the FAE1 polypeptide sequence

As a negative control, to demonstrate that the FAE1 gene sequence was unlikely to provide a useful search sequence in the identification of *C.elegans* sequences encoding for PUFA elongases, the GenBank databases (http://www.ncbi.nlm.nih.gov/Web/Search/index.html) were searched using the *Arabidopis* FAE1 polypeptide sequence to identify related genes or expressed sequence transcripts (ESTs). GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acid Research* (1998) **26**, 1-7). There are approximately 2,162,000,000 bases in 3,044,000 sequence records as of December 1998. The search was carried out using the BLAST2 (Basic Local Alignment Search Tool) algorithm (Altschul *et al.*, (1990) *J Mol Biol* **215**,403,410) Although a number

of plant ORFs and ESTs were reported as being related, no animal sequences were identified by this search, confirming the observation that FAE1 was unlikely to be a suitable candidate as a search template for PUFA elongases.

Database search using yeast ELO sequences

Using the three yeast fatty acid elongase sequences (ELO 1, 2, 3) as probes, a number of putative ORFs in the DNA of *C. elegans*-derived cosmid sequences which form the *C. elegans* genomic sequence database were identified. Moreover, an extensive and time-consuming search of a downloaded copy of the WormPep database (ftp://ftp.sanger.ac.uk./pub/databases/wormpep) using manual search strings in MSWord 6, identified a number of *C. elegans* ORFs which contained presumptive histidine boxes. Wormpep contains predicted proteins from the *Caenorhabditis elegans* genome sequence project, which is carried out jointly by the Sanger Centre in Cambridge, UK and Genome Sequencing Center in St Louis, USA. The current Wormpep database, Wormpep 16, contains 16,332 protein sequences (7,120,115 residues). Search strings used included [HXXHH], [HXXXHH], [QXXHH] and [YHH]. Comparison of the data from the two different searches indicated a small (<10) number of putative ORFs as candidate elongases. The histidine box motifs are shown in bold in SEQ ID 9 to 16.

Hydrophobicity plot analysis

Since the fatty acid elongase reaction is predicted to be carried out on the cytosolic face of the endomembrane system (Toke & Martin (1996), *supra*; Oh *et al* (1997), *supra)*, the putative *C. elegans* ORFs were examined for potential membrane spanning domains, via Kyte & Doolittle hydrophobicity plots *(J. Mol Biol,* (1982), **157**, 105-132). This revealed a number of ORFs with possible membrane-spanning domains, and also indicated a degree of similarity in the secondary-structure of a number of identified ORFs.

Screening for ER-retention signal sequences

The inventors postulated that since fatty acid elongases are expected to be endoplasmic reticulum (ER) membrane proteins, they might be expected to have peptide signals which are responsible for "ER-retention". In the case of ER membrane proteins, this signal often takes the form of a C-terminal motif [K-K-X_{2-3}-Stop], or similar variants thereof (Jackson *et*

al., (1990), *EMBO J.*, **9**, 3153-3162). Further sequence analysis of the *C. elegans* putative elongases revealed that 4 ORFs (F41H10.7, F41H10.8, F56H11.4, Y53F4B.c) had *C*-terminal motifs that exactly matched this search pattern, and that a further 2 ORFs (F11E6.5, C40H1.4) had related sequences. These sequence motifs are underlined in SEQ ID 9 to 13, 15 and 16.

Chromosome mapping

sub a6

Since the inventors had previously observed that *C.elegans* genes involved in the synthesis of PUFA may exist in tandem (for example the Δ5 and Δ6 desaturases required for AA and GLA synthesis, respectively, are < 1 kB apart on chromosome IV (Michaelson *et al.*, (1998), *FEBS Letts* **439**, 215-218), the positions of the putative *C. elegans* elongase ORFs were determined using the Sanger Centre's WebAce *C. elegans* server (http://www.sanger.ac.uk/Projects/C_elegans/webace_front_end.shtml).. This indicated that two pairs of putative elongases were in close proximity to each other on the *C. elegans* chromosome IV.

F41H10.7 and F41H10.8 were identified as being approximately 10 Kb apart on chromosome IV, and F56H11.3 and F56H11.4 were identified as being approximately 2 Kb apart on chromosome IV.

Putative *C. elegans* fatty acid elongases

The positions of the putative ORFs in the *C. elegans* genome are shown below i.e. chromosome number, and map position in centiMorgans, together with the GenBank database accession numbers.

The designations used employ the same method as used on the Sanger Centre's *C. elegans* database, i.e. ORF C40H1.4 is predicted coding sequence 4 on cosmid C40H1.

Elongase	Cosmid Sanger ID Code	GenBank Acc	Chromosome
A	C40H1.4	Z19154	III

B	D2024.3	U41011	IV, 7.68
C	F11E6.5	Z81058	IV, 18.8
D	F41H10.7*	U61954	IV, 29.8
E	F41H10.8*	U61954	IV, 29.8
F	F56H11.3#	Z68749	IV, 2.5
G	F56H11.4#	Z68749	IV, 2.5
H	Y53F4B.c	Z92860	II

* or # indicates genes in tandem

Comparison of *C. elegans* putative elongase ORFs with yeast genes:

Each of the three yeast ELO polypeptides were compared against all of the worm putative elongase translated ORF sequences, and then ranked in order of similarity (as measured by the BLAST score) (Altschul *et al* (1990), *supra*)

The results are shown below, with the ORF sequences ranked from most similar to least similar, and the BLAST scores are shown in brackets:

Yeast ELO1 (14 to 16 carbon fatty acid elongase)

G (262) > E (241) > D (225) > C (219) > A (216) > F (215) > H (197) > B (172)

Yeast ELO2 (24 carbon sphingolipid elongase)

E (231) > C (226) > G (189) > A (181) > F (166) > D (150) > H (141) > B (140)

Yeast ELO3 (24 to 26 sphingolipid elongase)

D (171) > G (163) > F (154) > A (152) > E (150) > C (131) > B (132) > H (128)

It is clear from the numeric values of the BLAST scores that the sequences are related, but the levels of homology are low. For comparison, the BLAST score for homology between two related worm proteins, the Δ5 and the Δ6 desaturase is in excess of 500.

Analysis of potential sphingolipid ancestry

Previously, the inventors had noted the similarities between the fatty acid Δ6 desaturase and sphingolipid desaturases in plants, and that the two distinct enzymes could have arisen from one ancestral gene. Moreover, it was considered likely that the sphingolipid desaturase predated the fatty acid desaturase, and may in fact have been the ancesteral progenitor. Therefore it is plausible that the next step in the arachidonic acid biosynthetic pathway has also evolved from the sphingolipid metabolic pathway. It is therefore considered highly significant that some of the *C. elegans* ORF putative elongases have similarity to sphingolipid enzymes. For this reason, these ORFs are considered to be very clear candidates for PUFA elongases. It has previously been considered that the *C. elegans* Δ5 and Δ6 fatty acid desaturases have evolved from 1 ancestral gene (Michaelson *et al.*, (1998), *FEBS Letts* **439**, 215-218). It is also significant that one pair of *C. elegans* putative elongase ORFs (F & G) genetically maps close to the Δ5/Δ6 fatty acid desaturase genes, with both gene pairs being located at the top end of chromosome IV.

Cosmid Sanger ID Code	GenBank Acc	Chromosome	Encoded Peptide
W08D2.4	Z70271	IV, 3.06	Δ6 fatty acid desaturase
T13F2.1	Z81122	IV, 3.06	Δ5 fatty acid desaturase

Cloning of Desaturase and Elongase Genes in Yeast Expression Vectors

Sub Q7

Putative elongases sequences F56H11.4 and F41H10.8 were cloned by PCR into the pYES2 vector (Invitrogen). A *C. elegans* mixed stage cDNA library was used as a PCR template. F56H11.4 was amplified using primers:

56h114.for 5'-GCGGGTACCATGGCTCAGCATCCGCTC-3' and;

56h114.rev 5'-GCGGGATCCTTAGTTGTTCTTCTTCTT-3'.

F41H10.8 was amplified using primers:

41h108.for 5'-GCGGGTACCATGCCACAGGGAGAAGTC-3' and;

41h108.rev 5'-GCGGGATCCTTATTCAATTTTTCTTTT-3'.

Amplified sequences were then restricted using *KpnI* and BamHI (underlined in the forward and reverse primers, respectively), purified using the Qiagen PCR purification kit, and ligated into a KpnI/BamHI cut pYes2 vector.

Sub Q8

An ORF encoding the *Mortierella alpina* Δ^5-fatty acid desaturase (Michaelson, L. V., *et al* (1998) *J. Biol. Chem.* **273**, 19055-19059) was amplified using primers:

Mad5.for 5'-GCGAATTCACCATGGGTACGGACCAAGGA-3' and;

Mad5.rev 5'-GCGGAGCTCCTACTCTTCCTTGGGACG-3',

and restricted using *Eco*RI and *Sac*I, gel purified as described and ligated into a *Eco*RI/*Sac*I cut pESC-TRP vector (Stratagene) to generate pESC/Δ^5.

An ORF encoding the borage Δ^6-fatty acid desaturase (Sayanova, O., *et al* (1997) *Proc. Natl. Acad. Sci USA* **94**, 4211-4216) was restricted from pGEM3 using *Bam*HI and *Xho*I and ligated into a *Bam*HI/*Xho*I cut pESC-TRP vector to generate pESC/Δ^6.

A double construct was also generated by ligating the *Bam*HI/*Xho*I borage Δ^6 insert into the pESC/Δ^5 construct described previously, generating pESC/(Δ^5,Δ^6).

Functional Characterisation in Yeast

Elongases and desaturase constructs were introduced in *Saccharomyces cerevisiae* W303-1A using a lithium acetate based method (Elble, R. (1992) *Biotechniques* **13**, 18-20) and expression of the transgenes was induced by addition of galactose to 2% (w/v) as described in Napier *et al* (Napier, J. A., *et al* (1998) *Biochem J* **330**, 611-614; Michaelson L. V., *supra*; Michaelson, L. V., (1998) *FEBS Letts* **439**, 215-218). Yeast transformants containing pYES2-derived constructs were grown on synthetic minimal media (SD, the composition of which is defined in Sherman, F (1991) *Methods in Enzymology* **194**, 3-21); synthetic minimal medium minus uracil; pESC-derived constructs were grown on SD minimal medium minus tryptophan. Co-transformed yeast (containing both pYES2 and pESC derivatives) were grown on SD minimal medium minus uracil and tryptophan. Prior to induction, cultures were grown in the presence of 2% raffinose and supplemented with 0.5 mM of the appropriate fatty acid substrate in the presence of 1% tergitol-(NP40) (Sigma). All cultures were then grown for a further 48-h unless indicated.

Fatty Acid Analysis

To identify the elongation reaction responsible for the synthesis of di-homo-γ-linolenic acid (DHGLA; 20:3$\Delta^{8,11,14}$) from GLA, this latter fatty acid was supplied as the (exogenous) substrate.

Lipids were extracted from transformed and control yeast by homogenisation in MeOH-$CHCl_3$ using a modification of the method of Bligh and Dyer (Dickenson & Lester (1999) *Biochim Biophys Acta* **1426**, 347-357). The resulting $CHCl_3$ phase was evaporated to dryness under nitrogen gas and the samples were transmethylated with 1M HCl in methanol at 80 °C for 1 hour. Fatty acid methyl esters (FAMES) were extracted in hexane and purified using a small column packed with Florisil. Analysis of FAMES was conducted using a Hewlett Packard 5880A Series Gas Chromatograph equipped with a 25M x 0.32mm RSL-500BP bonded capilliary column and a flame ionisation detector. Fatty acids were identified by comparison of retention times with FAME standards (Sigma)

separated on the same GC. Quantitation was carried out using peak height area integrals expressed as a total of all integrals (Bligh, E.G. & Dyer, W.J. (1959) *Can. J. Biochem. Physiol.* **37**, 911-917).

Total fatty acids extracted from yeast cultures were analysed by gas chromatography (GC) of methyl ester derivatives. Lipids were extracted, transmethylated and the fatty acid methyl esters (FAMEs) analysed as described by Sayanova *et al.*

Figure 11 shows chromatograms of fatty acid methyl esters from yeast transformed with the control (empty) plasmid pYES2 (Fig. 11A) or with ORF F56H11.4 in pYES2 (Fig. 11B). Exogenous substrate in the form of GLA was supplied to the cultures. Two novel peaks are observed in (B); these peaks (annotated as 20:3 and 18:1*) were identified (against known standards) as DHGLA and vaccenic acid, respectively. Detection was by flame ionisation.

One cDNA ORF tested in this manner displayed a high level of elongase activity on the GLA substrate, converting 44% to DHGLA. The identity of this elongation product was confirmed as DHGLA by comparison with a known standard (the standards used were known standards for either DHGLA, AA, EPA or VA from Sigma Chemicals, Ltd.), using GCMS analysis using a Kratos MS80RFA (Napier, J. A., *supra*; Michaelson, L. V., *supra*; Michaelson, L. V., *supra*). The deduced amino acid sequence of the functional elongase clone identified it as being encoded by the *C. elegans* gene F56H11.4, and comparison with the yeast *ELO* genes showed low homology confined to a few short amino acid motifs (see Fig. 10). Some similarity with a mouse gene Cig30 (Tvrdik, P., (1997) *J. Biol. Chem.* **272**, 31738-31746), which has been implicated in the recruitment of brown adipose tissue in liver tissue, was also observed, as well as a potential human homologue encoded by a gene located on chromosome 4q25, BAC 207d4. The most closely related *C. elegans* ORFs, F41H10.8 (U61954) and F56H11.3 (Z68749) are also shown, as is part of a related human gene present on chromosome IV (present on BAC clone B207d4; AC004050). The GenBank accession numbers are given for all sequences.

The range of fatty acids synthesised by *C. elegans* can potentially require a number of different elongation reactions (Tanaka, T., (1996) *Lipids* **31**, 1173-1178). The substrate-specificity of the F56H11.4 PUFA elongase was therefore determined using a

range of exogenously supplied fatty acids. This revealed that GLA is the major substrate, with a number of other fatty acids being elongated at a lower efficiency (see Table 1). Although most of these substrates are polyunsaturated fatty acids, it was unexpectedly observed that palmitoleoic acid (PA; 16:1 Δ^9) was also elongated by F56H11.4 to yield vaccenic acid (VA; 18:1 Δ^{11}). The biosynthetic pathway for VA is unclear, but the data indicate that it may be synthesised by elongation of Δ^9-monounsaturated 16C fatty acid.

The *C. elegans* PUFA elongase ORF F56H11.4 maps to the top of chromosome IV (at 4.32 cM) with a related sequence (F56H11.3; 51 % similarity) located 1,824bp downstream. Another *C. elegans* gene (F41H10.8) was also observed, which is present on chromosome IV, and which shows a slightly higher level (53%) of similarity to the PUFA elongase than F56H11.3 (see Fig. 10). However, when a PCR product encoding ORF F41H10.8 was expressed in yeast in a manner identical to that used for F56H11.4, the former failed to direct the elongation of any fatty acids, despite the provision of a range of substrates (see Table II).

In order to reconstitute the PUFA biosynthetic pathway in a heterologous system, the PUFA elongase F56H11.4 was expressed in yeast in conjunction with either the Δ^6- or Δ^5-fatty acid desaturases previously isolated and characterised by the inventor (Napier, J. A., *supra*; Michaelson, L. V., *supra*). Expression of the Δ^6-fatty acid desaturase and F56H11.4 was carried out in the presence of two different substrates (LA or ALA) while the Δ^5-fatty acid desaturase and the elongase were expressed in the presence of GLA only. This demonstrated that was possible to combine a desaturase and an elongase in yeast to generate significant amounts of a final "product" (see Table III). In the case of the elongase and the Δ^6-fatty acid desaturase, the reactions proved highly efficient with the production of 4.5% of DHGLA from the LA substrate. This resulted from 25% desaturation of the LA substrate to GLA, which was then elongated to DHGLA at a similar level of efficiency (18%). This is lower than the % conversion observed for GLA when supplied exogenously (see Table I), indicating that the *in vivo* production of substrates for elongation may be rate-limiting.

If ALA was used as a substrate, 27% of this was initially Δ^6-desaturated to yield octadecatetraenoic acid (OTA; 18:4 $\Delta^{6,9,12,15}$) but only 8% of was subsequently elongated to yield eicosatetraenoic acid (20:4 $\Delta^{8,11,14,17}$). Thus, the conversion efficiency of ALA to the final 20-carbon tetraenoic PUFA was only about 2.2%.

Since DHGLA is an *n-6* fatty acid, whilst the OTA-derived eicostetraenoic acid is an *n-3* type, this demonstrates that the elongase is capable of accepting both forms of essential fatty acid, albeit with different efficiencies. Verification was also provided that the 20C PUFAs synthesised in the yeast expression system were generated by the Δ^6-desaturation of 18C substrates which were subsequently elongated, as the Δ^6-desaturase showed no activity on 20:2 or 20:3 substrates (see Table III).

The combination of the Δ^5-desaturase and the elongase also demonstrated that these two enzymes could work in tandem, although the efficiency of this overall conversion was lower (3.3% AA from GLA) which was due to the previously observed low activity of the Δ^5-desaturase enzyme itself (Michaelson, L. V., *supra*; Michaelson, L. V., *supra*). Thus, although nearly 45% of the GLA substrate was elongated to DHGLA, only 7.5% of this was then desaturated to AA (see Table III).

Finally, the production of either AA or eicosapentanoic acid (EPA; 20:5$\Delta^{5,8,11,14,17}$) in yeast from dienoic or trienoic 18 carbon substrates was achieved via expression of all three enzymes (the two desaturases and the F56H11.4 PUFA elongase) simultaneously. As shown in Table IV, small but significant amounts of AA were produced when the yeast was supplied with the 18C dienoic fatty acid LA.

GC-Mass Spectroscopy (MS) Analysis

Peak identification and confirmation were carried out by GC-MS using a Kratos MS80RFA using known standards (Sigma). The identity of this 20C PUFA was verified by GCMS, indicating that the conversion efficiency from LA was 0.65%. When ALA was used as a substrate, 12.5% of the (Δ^6-desaturated and elongated) eicosatetraenoic *n-3* fatty acid was Δ^5-desaturated, resulting in a total conversion of 0.3% of the ALA substrate to EPA (the identity of EPA was confirmed by GCMS).

Expression of *C. elegans* elongase in plants

In order to express *C. elegans* elongase in plants, the following protocol is an example of a process which can be used to create the transgenic plants. *C. elegans* ORF sequence can be subcloned into a plant expression vector pJD330, which comprises a viral 35S promoter, and a Nos terminator. The resulting cassette or promoter/coding sequence/terminator can then be subcloned into the plant binary transformation vector pBin 19, and the resulting plasmid introduced into *Agrobacterium tumefaciens.* This *Agrobacterium* strain can then be used to transform Arabidopsis by the vacuum-infiltration of inflorescences, and the seeds harvested and plated onto selective media containing kanamycin. Since pBin 19 confers resistance to this antibotic, only transformed plant material will grow. Resistant lines can therefore be identified and self-fertilized to produce homozygous material. Leaf material can then be analyzed for expression of *C. elegans* elongase.

Fatty acid methyl ester analysis can be carried out as previously described.

Table I

mole% Fatty Acids

F56H11.4

ORF	-											
	(Control)											
Substrate	-	-		GLA		LA		ALA		EPA	
Induction	-	+	-	+	-	+	-	+	-	+	-
16:0	17.5 ± 3.3	19.9 ± 3.5	20.5 ± 4.1	27.7 ± 1.4	29.8 ± 0.2	22.9 ± 1.5	23.9 ± 1.0	19.1 ± 0.7	20.2 ± 1.1	23.4 ± 0.2	24.2 ± 1.0
16:1	53.2 ± 7.2	40.9 ± 3.1	49.4 ± 3.2	32.5 ± 4.4	34.4 ± 1.8	21.2 ± 2.2	24.8 ± 4.9	18.1 ± 1.5	19.6 ± 2.5	26.9 ± 0.7	34.5 ± 1.5
18:0	4.5 ± 0.7	4.7 ± 0.9	4.9 ± 0.5	5.6 ± 0.5	5.6 ± 0.3	5.1 ± 0.3	4.4 ± 0.1	5.0 ± 0.3	4.5 ± 0.1	5.3 ± 0.2	5.4 ± 0.2
18:1	24.8 ±3.9	24.9 ± 1.4	25.2 ± 2.3	16.9 ± 0.9	16.1 ± 0.3	11.2 ± 2.4	10.7 ± 1.5	10.1 ± 1.1	9.9 ± 1.2	15.4 ± 0.4	17.8 ± 1.1
18:1*	-	9.6 ± 0.6	-	3.9 ± 0.6	-	3.2 ± 0.6	-	3.1 ± 0.4	-	6.2 ± 0.3	-
LA	-	-	-	-	-	34.4 ± 4.2	36.2 ± 5.6	-	-	-	-
ALA	-	-	-	-	-	-	-	43.1 ± 3.9	45.8 ± 4.8	-	-
GLA	-	-	-	7.5 ± 1.2	14.0 ± 0.3	-	-	-	-	-	-
20:2	-	-	-	-	-	2.0 ± 0.9	-	-	-	-	-
DHGLA	-	-	-	5.8 ± 0.9	-	-	-	-	-	-	-
20:3	-	-	-	-	-	-	-	1.5 ± 0.1	-	-	-
EPA	-	-	-	-	-	-	-	-	-	22.8 ± 0.7	20.1 ± 2.4
% Elongated											
GLA	-	-	-	44	-	-	-	-	-	-	-
LA	-	-	-	-	-	5.5	-	-	-	-	-
ALA	-	-	-	-	-	-	-	3.4	-	-	-
EPA	-	-	-	-	-	-	-	-	-	0	-

Table II

mole% Fatty Acids

ORF	F41H10.8									
Substrate	-		GLA		LA		ALA		EPA	
Induction	+	-	+	-	+	-	+	-	+	-
16:0	19.0 ± 0.9	19.3 ± 0.2	28.1 ± 0.6	28.0 ± 0.9	23.9 ± 0.7	24.4 ± 0.2	22.8 ± 0.2	23.4 ± 0.2	23.0 ± 0.6	23.7 ± 0.9
16:1	50.9 ± 0.7	50.8 ± 0.6	33.5 ± 2.2	35.5 ± 1.5	22.4 ± 2.1	23.6 ± 0.3	17.6 ± 0.2	15.8 ± 0.9	34.7 ± 3.6	32.2 ± 3.2
18:0	4.2 ± 0.1	5.1 ± 0.1	5.3 ± 0.1	5.6 ± 0.1	5.1 ± 0.2	5.8 ± 0.1	5.4 ± 0.3	5.9 ± 0.1	4.8 ± 0.7	5.1 ± 0.3
18:1	24.5 ± 1.3	24.9 ± 0.5	16.2 ± 1.4	17.1 ± 1.0	9.1 ± 0.3	10.1 ± 0.2	7.8 ± 0.1	9.5 ± 0.6	15.3 ± 2.5	15.3 ± 1.8
18:1*	ND	-	ND	-	ND	-	ND	-	ND	-
LA	-	-	-	-	39.5 ±0.6	36.1 ± 0.4	-	-	-	-
ALA	-	-	-	-	-	-	46.4 ± 0.5	45.4 ± 1.3	-	-
GLA	-	-	14.3 ± 1.6	14.2 ± 0.6	-	-	-	-	-	-
20:2	-	-	-	-	ND	-	-	-	-	-
DHGLA	-	-	ND	-	-	-	-	-	-	-
20:3	-	-	-	-	-	-	ND	-	22.3 ± 2.8	23.8 ± 2.2
EPA	-	-	-	-	-	-	-	-		
% Elongated										
GLA	-	-	0	-	-	-	-	-	-	-
LA	-	-	-	-	0	-	-	-	-	-
ALA	-	-	-	-	-	-	0	-	-	-
EPA	-	-	-	-	-	-	-	-	0	-

Table III

mole% Fatty Acids

Construct	Δ^6		F56H11.4 + Δ^6				F56H11.4 + Δ^5	
			LA		ALA		GLA	
Substrate	20:2	20:3						
Induction	+	-	+	-	+	-	+	-
16:0	24.7 ± 1.3	25.2 ± 1.5	18.7 ± 0.6	23.7 ± 0.5	17.4 ± 0.7	21.0 ± 1.3	27.9 ± 4.2	29.8 ± 3.8
16:1	46.0 ± 2.8	43.7 ± 3.7	18.9 ± 1.2	24.6 ± 0.7	5.3 ± 0.6	9.1 ± 0.9	24.6 ± 3.4	25.1 ± 3.2
16:2	5.2 ± 1.2	4.1 ± 1.4	0.6 ± 0.1	-	0.4 ± 0.1	-	-	-
18:0	4.8 ± 0.4	5.1 ± 0.4	4.0 ± 0.3	5.1 ± 0.1	6.2 ± 0.7	5.4 ± 0.2	5.6 ± 0.8	5.4 ± 0.7
18:1	15.3 ± 1.1	16.1 ± 1.2	12.2 ± 1.4	11.2 ± 0.4	5.7 ± 0.8	6.0 ± 0.4	12.7 ± 2.9	13.0 ± 2.5
18:1*	-	-	7.7 ± 0.7	-	2.6 ± 0.3	-	2.9 ± 0.9	-
LA	-	-	25.0 ± 3.2	35.4 ± 2.1	-	-	-	-
ALA	-	-	-	-	42.3 ± 3.3	58.5 ± 4.7	-	-
GLA	-	-	7.9 ± 2.2	-	-	-	13.2 ± 3.6	19.2 ± 3.5
OTA	-	-	-	-	15.3 ± 1.8	-	-	-
20:2	4.0 ± 0.3	-	3.3 ± 0.5	-	-	-	-	-
DHGLA	-	-	1.7 ± 0.2	-	-	-	9.8 ± 1.8	-
20:3	-	5.8 ± 0.5	-	-	3.4 ± 0.4	-	-	-
AA	-	-	-	-	-	-	0.8 ± 0.2	-
20:4	-	-	-	-	1.4 ± 0.2	-	-	-
EPA	-	-	-	-	-	-	-	-
% Elongated								
GLA	-	-	17.7	-	-	-	44.5	-
OTA	-	-	-	-	8.4	-	-	-
LA	-	-	8.7	-	-	-	-	-
ALA	-	-	-	-	5.4	-	-	-

SEQ ID1

C40H1.4

atggagcttgccgagttctggaatgatctcaacaccttcaccatctacggaccgaatcac
acagatatgaccacaaaatacaaatattcatatcacttcccaggtgaacaggtggcggat
ccgcagtattggacgattttattccagaaatattggtatcattcgatcacaatatcagtt
ctttatttcattttaattaaggtgattcaaaagtttatggagaatcgaaaaccattcact
ttgaaatacccattgattctttggaatggagctcttgcagcattcagtataattgccaca
ttgcggttctctattgatcctctacgatcactatatgctgaaggattctacaaaactctg
tgctattcgtgtaatccaactgatgtggctgcattttggagctttgcattcgctctttcc
aagattgttgaacttggagacactatgttcattattttgagaaaacggccattgatcttt
ttacactactatcatcatgcagcagtgttaatctacactgtccattctggtgccgagcat
actgcagctggtcgtttctacatcctaatgaactacttcgcacattctctcatgtatact
tactacacagtttctgccatgggatacagattaccgaaatgggtatcaatgactgtcaca
actgttcaaacaactcaaatgttagctggagtcggaataacttggatggtgtacaaagtg
aaaactgaatacaagcttccttgtcaacaatccgtagccaatttgtatctcgcattcgtc
atctatgtcacatttgccattcttttcattcaattcttcgtcaaggcatacattatcaag
tcgtcgaagaagtcgaaatcggtgaagaacgaataa

SEQ ID2

D2024.3

atggcaaaatacgactacaatccgaagtatgggttagaaaattacagcatattccttccc
tttgagacatcttttgatgcatttcgatcgacaacatggatgcaaaatcactggtatcaa
tcaattacagcatctgtcgtgtatgtagccgtcatttttacaggaaagaaggtggttctc
atctacaaaaaatcacgagttattacttttgagtctagccttcagaatgcaattaagaat
cgaaaccgaaaatcacttaatagttctcaaatgtttcagattatggaaaagtacaagccc
ttccaactggacacaccactcttcgtctggaattcatttttagccattttctcaattctc
gggttcctccgaatgacacctgaatttgtatggagttggtcagcagaaggaaactcattc
aaatattcaatttgtcattcatcttatgctcaaggagtcactggtttctggactgaacaa
ttcgcaatgagcaaacttttcgagctcatcgacacaatcttcatcgttcttcgtaaacgt
ccactcatcttccttcactggtatcatcatgtaactgttatgatctacacatggcacgcg
tacaaggatcacactgcatcaggacggtggttcatttggatgaattatggagttcatgct
cttatgtattcctactatgctcttcgttctctgaaattccgtcttccaaaacaaatggca
atggttgttactactctccaacttgctcaaatggttatgggagtaatcatcggagtcact
gtctaccgtatcaagtcatcgggtgaatactgccaacagacatgggacaatttgggatta
tgctttggagtttatttcacatatttccttcttttcgccaacttcttctaccatgcatat
gttaagaaaaacaaccgtacagtaaattatgaaaataattcaaaaaatttccccgatctc
gttttaatttacctgagaaaaaaggtttcaagaaaatcgaaaaatcggcaatgttcagaa
aataattataaaattcaattttcatcaaattttgttaatgttgatggaaaaaaacataag
aaaacatatgaacttattcttccaagaagaaaaatgaccacaattttaacttttctattt
ggaaaaaatcgaatttttttcgaaatatcagaaaaatcgaaaaaacatttcgattcctgtt
gatttcgaaattctggagccaaaagaagatatcaatgctaacatcgctgagccatccatc
acaacgaggtccgccgccgcacgaagaaaagttcaaaaagctgattag

SEQ ID3

F11E6.5

atggcagcagcacaaacaagtccagcagccacgctcgtcgatgttttgacaaaaccatgg
agtctggatcagactgattcttacatgtctacatttgtaccattatcctataaaatcatg
attggttatctcgtcaccatctacttcgggcaaaaattaatggctcacagaaaaccattc
gatctccaaaatacacttgctctctggaacttcgggttttcactgttctcgggaatcgcc
gcctataagcttattccagaactattcggagttttcatgaaggacgggtttgtcgcttcc
tactgtcaaaacgagaactactacaccgatgcatcaactggattctggggctgggccttt
gtgatgtcgaaagctccagaactaggggatactatgttcttggtccttcgtaaaaaacca
gttatcttcatgcactggtatcatcatgccctcacatttgtctacgcagtagtcacatac
tctgagcatcaggcatgggctcgttggtctttggctctcaaccttgccgtccacactgtt
atgtatttctacttcgccgttcgcgccttgaacatccaaactccacgcccagtggcaaag
ttcatcactactattcaaattgtccaatttgtcatctcatgctacatttttgggcatttg
gtattcattaagtctgctgattctgttcctggttgcgctgttagctggaatgtgctatcg
atcggaggactcatgtacatcagttatttgttccttttgccaagttcttctacaaggcc
tacattcaaaaacgctcaccaaccaaaaccagcaagcaggagtag

SEQ ID4

F41H10.7

atgtcatcggacgatcgtggcactagaaccttcaagatgatggatcaaattcttggaaca
aacttcacttatgaaggtgccaaagaagttgctcgaggccttgaaggtttctcagcaaag
cttgccgtcggatatattgccactatttttggactgaaatattatatgaaagaccgaaaa
gccttcgatctcagtactccattaaacatttggaatggtattctttcgacattcagctta
ttgggattcttattcacttttcctactttgttatcagttatcagaaaggatggatttagt
cacacctattcccatgtctctgagctttacactgacagtacctctggatattggatcttc
ctttgggttatctcaaagattccggaacttttggatacagtattcattgttcttcgcaag
agaccacttattttcatgcactggtaccatcacgcattgaccggttactatgctcttgtc
tgctaccatgaggatgctgtccatatggtttgggttgtatggatgaattatattattcat
gcattcatgtatggatactatcttctgaaatctctgaaagttccaattccaccatcagtt
gctcaagcaatcaccacatctcaaatggttcaattcgcagttgccattttcgcacaagtt
catgtttcctataaacactatgttgagggagttgaaggattagcctactcgttcagagga
acagctatcggatttttcatgcttactacctacttctatctatggattcaattctacaaa
gagcactatcttaagaatggaggcaaaaagtacaatttggcaaaggatcaggcaaaaact
caaacaaagaaggctaactaa

SEQ ID5

F41H10.8 (ce477)

atgccacagg gagaagtctc attctttgag gtgctgacaa ctgctccatt
cagtcatgag ctctcaaaaa agcatattgc acagactcag tatgctgctt
tctggatctc aatggcatat gttgtcgtta tttttgggct caaggctgtc
atgacaaacc gaaaaccatt tgatctcacg ggaccactga atctctggaa
tgcgggtctt gctatttttct caactctcgg atcacttgcc actacatttg
gacttctcca cgagttcttc agccgtggat ttttcgaatc ttacattcac
atcggagact tttataatgg actttctgga atgttcacat ggcttttcgt

tctctcaaaa gttgctgaat tcggagatac acttttttatt attcttcgta
aaaagccatt gatgttcctt cattggtatc atcatgtgct tacaatgaat
tatgctttta tgtcatttga agctaatttg ggatttaata cttggattac
atggatgaat ttctcagttc actcaattat gtatggatat tatatgcttc
gttcttttgg tgtcaaggtt ccagcatgga ttgccaagaa tattacaaca
atgcaaattc ttcaattcgt tattactcat ttcattcttt tccacgttgg
atatttggca gttactggac aatctgttga ctcaactcca ggatattatt
ggttctgcct tctcatggaa atctcttatg tcgttctgtt cggaaacttc
tactatcaat catacatcaa gggaggtggc aagaagttta atgcagagaa
gaagactgaa aagaaaattg aataa

SEQ ID6

F56H11.3

atgtatttgaattatttcgcgacggaaatcttccatcgtagtgcggtttgtgaaacagaa
gcttgtcgctcgtcaaaaataatgattgctgacgtgttcaaatggaaattcgatgcaaac
gaattgtggagtcttttaacgaatcaggatgaagttttcccgcatattagagcacggcga
ttcattcaagaacattttggtctattcgtccagatggcaattgcatatgtcattttggtg
ttctcaatcaaaaggttcatgagggatcgtgaaccatttcaactcaccacagctcttcgt
ctctggaacttcttcctctccgtcttctcaatttatggttcctggacaatgtttccattt
atggttcaacaaataagactttatggtctctacggatgtggatgcgaagcactttcaaac
cttccgagtcaagcagaatattggcttttcctgacgatcttgtccaaagctgtggagttt
gttgatacatttttcttggttctccggaaaaaaccactcatcttcctacactggtatcat
catatggcaacatttgtcttcttctgcagtaattacccgactccatcgtcacaatcacgc
gtcggagttatcgtcaacctgttcgtgcatgccttcatgtacccatactatttcacccga
tcaatgaacatcaaagttcctgcgaaaatttcaatggctgttacagttcttcaattgact
caattcatgtgctttatctatggatgtactctcatgtactactcgttggccactaatcag
gcacgataccctcaaatacacctgcgacactccaatgtttgtcctacactctacatttg
ctttga

SEQ ID7

F56H11.4 (Ce 166)

atg gctcagcatc cgctcgttca acggcttctc gatgtcaaat tcgacacgaa
acgatttgtg gctattgcta ctcatgggcc aaagaatttc cctgacgcaga
aggtcgcaa gttctttgct gatcactttg atgttactat tcaggcttcaa
tcctgtacat
ggtcgttgtg ttcggaacaa aatggttcat gcgtaatcgt caaccattcc
aattgactat tccactcaac atctggaatttcatcctcgc cgcattttcc
atcgcaggag ctgtcaaaat gaccccagag ttctttggaa ccattgccaa
caaaggaatt gtcgatcctactgc aaagtgtttg atttcacgaa aggagagaat
ggatactgggt gtggctctt catggcttcc aaacttttcg aacttgttga
caccatcttc ttggttctccgtaaacgtcc actcatgttc cttcactggt
atcaccatat tctcaccatg atctacgcctggtactctca tccattgacc
ccaggattca acagatacgg aatttatctt aactttgtcg
tccacgcctt catgtactct tactacttcc ttcgctcgat gaagattcgc gtgc
caggattcatcgccca agctatcaca tctcttcaaa tcgttcaatt catcatctc

t tgcgccgttcttgctcatct tggttatctc atgcacttca ccaat
gccaactgt gatttcgagc catcagtatt caagctcgca
gttttcatgg acacaacata cttggctctt ttcgtcaact tcttcctcca
atcatatgtt
ctccgcggag gaaaagacaa gtacaaggca gtgccaaaga agaagaacaa ctaa

SEQ ID8

Y53F4B.c

atgtcggccg aagtgtccga acgattcaaa gtttggacag gaaacaatga
gaccatcatc tattccccat tcgagtacga ttccacgttg ctcatcgagt
catgtcggtg tacttatcag ctgcttatat tattgcgaca aatttattac
agagatatat ggagtcacgg aaacctaaaa cttttactag catggaacgg
ttttttggca gtgttcagta ttatgggtac atggagattt ggaatcgaat
tctacgatgc tgttttcaga agaggcttca tcgattcgat ctgcctggct
gtaaatccac gttcaccgtc cgcattctgg gcatgcatgt tcgctctatc
gaaaatcgcc gagtttgggg acacgatgtt cttggtgctg aggaaacggc
cggttatatt ccttcactgg tatcatcacg ctgttgttct gatcctttct
tggcatgctg caatcgaact cacagctcca ggacgctggt ttatttttat
gaactatttg gtgcattcaa taatgtatac atactacgca ataacatcaa
tcggctatcg tcttcccaaa atcgtttcaa tgactgttac attccttcaa
actcttcaaa tgctcattgg tgtcagcatt tcttgcattg tgctttattt
gaagcttaat ggagagatgt gccaacaatc ctacgacaat ctggcgttga
gcttcggaat ctacgcctca ttcctggtgc tattctccag tttcttcaac
aatgcatatt tggtaaaaaa ggacaagaaa cccgatgtga agaaggatta
a

SEQ ID9

A

1 MELAEFWNDL NTFTIYGPNH TDMTTKYKYS YHFPGEQVAD PQYWTILFQK
51 YWYHSITISV LYFILIKVIQ KFMENRKPFT LKYPLILWNG ALAAFSIIAT
101 LRFSIDPLRS LYAEGFYKTL CYSCNPTDVA AFWSFAFALS KIVELGDTMF
151 IILRKRPLIF L**HYYHH**AAVL IYTVHSGAEH TAAGRFYILM NYFAHSLMYT
201 YYTVSAMGYR LPKWVSMTVT TVQTTQMLAG VGITWMVYKV KTEYKLPCQQ
251 SVANLYLAFV IYVTFAILFI QFFVKAYIIK SSKKSKSVKN E*

SEQ ID10

B

1 MAKYDYNPKY GLENYSIFLP FETSFDAFRS TTWMQNHWYQ SITASVVYVA
51 VIFTGKKVVL IYKKSRVITF ESSLQNAIKN RNRKSLNSSQ MFQIMEKYKP
101 FQLDTPLFVW NSFLAIFSIL GFLRMTPEFV WSWSAEGNSF KYSICHSSYA
151 QGVTGFWTEQ FAMSKLFELI DTIFIVLRKR PLIFL**HWYHH** VTVMIYTWHA
201 YKDHTASGRW FIWMNYGVHA LMYSYYALRS LKFRLPKQMA MVVTTLQLAQ

```
251  MVMGVIIGVT VYRIKSSGEY CQQTWDNLGL CFGVYFTYFL LFANFFYHAY
301  VKKNNRTVNY ENNSKNFPDL VLIYLRKKVS RKSKNRQCSE NNYKIQFSSN
351  FVNVDGKKHK KTYELILPRR KMTTILTFLF GKNRIFSKYQ KNRKNISIPV
401  DFEILEPKED INANIAEPSI TTRSAAARRK VQKAD*
```

SEQ ID11

C

```
1    MAAAQTSPAA TLVDVLTKPW SLDQTDSYMS TFVPLSYKIM IGYLVTIYFG
51   QKLMAHRKPF DLQNTLALWN FGFSLFSGIA AYKLIPELFG VFMKDGFVAS
101  YCQNENYYTD ASTGFWGWAF VMSKAPELGD TMFLVLRKKP VIFMHWYHHA
151  LTFVYAVVTY SEHQAWARWS LALNLAVHTV MYFYFAVRAL NIQTPRPVAK
201  FITTIQIVQF VISCYIFGHL VFIKSADSVP GCAVSWNVLS IGGLMYISYL
251  FLFAKFFYKA YIQKRSPTKT SKQE*
```

SEQ ID12

D

```
1    MSSDDRGTRT FKMMDQILGT NFTYEGAKEV ARGLEGFSAK LAVGYIATIF
51   GLKYYMKDRK AFDLSTPLNI WNGILSTFSL LGFLFTFPTL LSVIRKDGFS
101  HTYSHVSELY TDSTSGYWIF LWVISKIPEL LDTVFIVLRK RPLIFMHWYH
151  HALTGYYALV CYHEDAVHMV WVVWMNYIIH AFMYGYYLLK SLKVPIPPSV
201  AQAITTSQMV QFAVAIFAQV HVSYKHYVEG VEGLAYSFRG TAIGFFMLTT
251  YFYLWIQFYK EHYLKNGGKK YNLAKDQAKT QTKKAN*
```

SEQ ID13

E

```
MPQGEVSFFE VLTTAPFSHE LSKKHIAQTQ YAAFWISMAY VVVIFGLKAV
MTNRKPFDLT GPLNLWNAGL AIFSTLGSLA TTFGLLHEFF SRGFFESYIH
IGDFYNGLSG MFTWLFVLSK VAEFGDTLFI ILRKKPLMFL HWYHHVLTMN
YAFMSFEANL GFNTWITWMN FSVHSIMYGY YMLRSFGVKV PAWIAKNITT
MQILQFVITH FILFHVGYLA VTGQSVDSTP GYYWFCLLME ISYVVLFGNF
YYQSYIKGGG KKFNAEKKTE KKIE*
```

SEQ ID14

F

```
1    MYLNYFATEI FHRSAVCETE ACRSSKIMIA DVFKWKFDAN ELWSLLTNQD
51   EVFPHIRARR FIQEHFGLFV QMAIAYVILV FSIKRFMRDR EPFQLTTALR
101  LWNFFLSVFS IYGSWTMFPF MVQQIRLYGL YGCGCEALSN LPSQAEYWLF
```

151 LTILSKAVEF VDTFFLVLRK KPLIFL**HWYH** **H**MATFVFFCS NYPTPSSQSR
201 VGVIVNLFVH AFMYPYYFTR SMNIKVPAKI SMAVTVLQLT QFMCFIYGCT
251 LMYYSLATNQ ARYPSNTPAT LQCLSYTLHL L* -

SEQ ID15

G

MAQHPLVQRL LDVKFDTKRF VAIATHGPKN FPDAEGRKFF ADHFDVTIQA
SILYMVVVFG TKWFMRNRQP FQLTIPLNIW NFILAAFSIA GAVKMTPEFF
GTIANKGIVA SYCKVFDFTK GENGYWVWLF MASKLFELVD TIFLVLRKRP
LMFL**HWYHH**I LTMIYAWYSH PLTPGFNRYG IYLNFVVHAF MYSYYFLRSM
KIRVPGFIAQ AITSLQIVQF IISCAVLAHL GYLMHFTNAN CDFEPSVFKL
AVFMDTTYLA LFVNFFLQSY VLRGGKDKYK AVPK<u>KKNN*</u>

SEQ ID16

H

MSAEVSERFKVWTGNNETIIYSPFEYDSTLLIESCRCTYQLLILLRQI
YYRDIWSHGNLKACDXLLLAWNGFLAVFSIMGTWRFGIEFYDAVFRXG
FIXSICLAVNPRSPSAFWACMFALSKIAEFGDTMFLVLRKRPVIFL**HWYHH**
AVVLILSWHAAIELTAPGRWFIFMNYLVHSIMYTYYAITSIGYRXPKIVSMT
VTFLQTLQMLIGVSISCIVLYLKLNGEMCQQSYDNLALSFGIYASFLVLSSFF
NNAYLVKKDKKPDV<u>KKD*</u>